Semantic Annotation and Retrieval of Digital Photos

Mathias Lux, Jutta Becker and Harald Krottmaier

Know-Center, Inffeldgasse 16c, A-8010 Graz, Austria
[mlux, jbecker]@know-center.at
Institute for Information Processing and Computer Supported new Media (IICM),
Graz, University of Technology, Inffeldgasse 16c, A-8010 Graz, Austria
hkrott@iicm.edu

Abstract. Nowadays research and development activities are accompanied by an increasing focus on future user needs in the field of multimedia retrieval. The fast growing of multimedia data repositories is an undeniable fact, so specialized tools allowing storage, indexing and retrieval of multimedia content have to be developed, and in addition easy-to-use content exchange is needed. The transition from text to photo retrieval raises the necessity of generating, storing and visualizing additional meta-information about the content to allow semantic retrieval. "Caliph & Emir" [5], a pair of prototypes allowing semantic annotation and retrieval of digital photos, are presented as a possible new way of handling semantics in descriptions of multimedia data.

Introduction

The evolution of digital information repositories produces more and more specialized requirements towards intelligent information retrieval. Base for interdisciplinary future developments are overall agreed standards and standardized methods. We chose following scenario to define the requirements, although a more generalized class of problems can be solved with our solution.

A user of a digital camera produces lots of images throughout a year and saves them to his local hard disk. The number of taken photos soon exceeds the critical mass for being manageable without specialized tools. To overcome this problem most people create an intuitive structure for storing their personal image library. They create folders for images taken in the same context or at the same time, for example "Photos from the I-Know '02 conference" or "Birthday party of uncle Adam". Nevertheless this does not enable a user to find a photo that shows a certain person, object, expresses a specific idea or feeling or even shows a scene, like "a picture showing a barking dog", when needed. Some file formats like TIFF and JPEG allow users to enrich the visual information with structured textual descriptions, but they only offer limited possibilities like adding keywords or adding a textual description.

In this paper the realization of Caliph & Emir, the software prototypes developed for supporting semantic annotation and retrieval of digital photos, is discussed.

Existing metadata standards for describing multimedia

The standard being used to define the way of handling the metadata has to be a lot more powerful than EXIF or for instance Dublin Core. DC only defines 15 core qualifiers, which can be understood as metadata tags, which can be filled by the user. A combination of Dublin Core and adapted Resource Description Framework structures, called RDF, would at least permit a structured storage of graphs similar to the above example and a quality rating, although content based image retrieval would not be supported.

MPEG-7 [6], the "Multimedia Content Description Interface", is an international standard, defined by the Moving Pictures Expert Group, which specifies a metadata system for describing multimedia content. MPEG-7 defines structures, called descriptors, needed for annotation of audio, video and image data and a standardized way to create new descriptors. A descriptor consists of either values or other descriptors; the format for defining those descriptors is W3C's XML Schema.

MPEG-7 matches the above requirements [1] and because MPEG-7 is a standard for metadata-based description of multimedia, it is interchangeable between various programs from different vendors. Besides it is possible to integrate existing standards, for instance Dublin Core [11], in MPEG-7.

To the best of our knowledge at this time no MPEG-7 based tools for managing digital photo repositories are available.

Realization of Caliph & Emir

"Caliph", the "Common And Lightweight Interactive PHoto annotation", was designed for supporting the user in the time consuming task of annotating photos, for retrieval of the annotated photos "Emir", the "Experimental Metadata based Image Retrieval", was developed. Both, mainly autonomous, prototypes Caliph and Emir were implemented using Suns Java SDK 1.4 while as runtime environment the versions JRE 1.4 and higher are supported.

Annotation with Caliph

Central part of Caliph is the so called "semantic description panel". It allows the user to define semantic objects like agents, places, events and times which are saved on exit for reusing them the next time starting Caliph. These semantic objects can also be imported from an existing MPEG-7 file to allow exchange of objects between users and editing and creating those objects in a user preferred tool. Semantic objects can be used for creating the description by dragging and dropping them onto the blue panel with the mouse, shown in figure 1. As once the objects exist, they can be reused if some pictures or series have the same context. This is especially true for objects representing persons, animals and places like the relatives, colleagues, friends, favorite pets or places like "at home" or "at work".
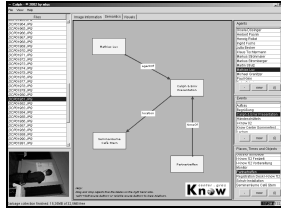
Fig. 1. Creating a semantic description using Caliph by drawing a graph as abstraction of the semantics

After dropping all the needed objects onto the blue panel the user can interconnect these objects by drawing relations (visualized by arrows) between them using the middle mouse button. The directed graph, which is generated through these user interactions with Caliph, can be saved as part of an MPEG-7 description.

Retrieval with Emir

Emir gives the user the ability to retrieve annotated photos. Due to the fact, that this is experimental software the retrieval mechanism is file system based. All MPEG-7 documents found by Emir in a specified directory and in further subdirectories are searched.

Besides searching for annotated images using keywords, Emir offers content based image retrieval based on the MPEG-7 ScalableColor descriptor and the ColorLayout descriptor. A detailed discussion of the visual descriptors can be found in [7].

The component of most interest is the panel offering a search mechanism for searching semantic descriptions. This component allows the user to define a graph with minimum one to maximum three nodes and two possible relations, where an asterisk is used as wildcard. A search graph which only contains one node with a word defining this node will return each MPEG-7 document wherein a semantic object containing the specified word is found. If two or three nodes and one or two relations are used to define the search graph, the repository of MPEG-7 documents is filtered by the terms defined as objects or relations and the remaining documents are searched for a structural match with the search graph.

Conclusion

MPEG-7 matches many of the current requirements for a metadata standard for usage in a personal digital photo library and it defines a lot more useful descriptors, which could be integrated as features in such libraries. In addition it is not only a standard for describing the content of images, but it also defines ways to annotate video and audio documents and it is prepared for general usage with multimedia data and cross-media retrieval.

The directed graph for describing semantics proves sufficient for usage with digital photos. It offers a lot more options than simple "tagging" mechanisms of photos like are supported in Adobe Photoshop Album [8] or the Photofinder project [9].

Some work has to be done on retrieval mechanism because although MPEG-7 defines similarity measurement for low-level content based descriptors [7] it fails to define those measurement methods for calculating the similarity of two semantic graphs.

Another problem is that different users produce different descriptions for the same media instance. Also MPEG-7 defines an inverse for each semantic relation, as a result a user can choose either he takes a relation or its inverse to create the description.

Acknowledgement

References

[1] Semantics of Multimedia in MPEG-7, Benitez Ana B. et al, 2002, International Conference on Image Processing 2002, Proceedings Volume: 1, Page(s): 137 -140

[2] Associative and Spatial Relationships in Thesaurus-Based Retrieval, Harith Alani et al, ECDL 2000, Proceedings, pp. 45–58

[3] Image Description and Retrieval Using MPEG-7 Shape Descriptors, Carla Zibreira, Fernando Pereira, ECDL 2000, Proceedings, pp. 332–335

[4] Everything You Want to Know About MPEG-7, F. Nack, A. Lindsay, Part 1, IEEE Multimedia, 6 (3), Juli-September 1999, 65-77

[5] Download Caliph & Emir, Know-Center Downloads & Demos, last visited 28.02.2003, http://www.know-center.at/en/divisions/div3demos.htm

[6] Multimedia Content Description Interface — Part 5: Multimedia Description Schemes, MPEG, 23.10.2001

[7] Multimedia content description interface – Part 3 Visual, MPEG, Juli 2001

[8] Adobe Photoshop Album, Adobe Inc, last visited 19.03.2003, http://www.adobe.com

[9] Photofinder project, University of Maryland Human Computer Interaction Laboratory, last visited 24.11.2002, http://www.cs.umd.edu/hcil/photolib/

[10] MPEG and its Relevance for Content-based Multimedia Retrieval, Haas Werner, Mayer Harald, J.UCS Vol 7 Issue 6, 10.7.2001, http://www.jucs.org/jucs_7_6/mpeg_and_its_relevance

[11] Proposal for the Integration of DublinCore and MPEG-7, Hunter, Jane, October 2000, last visited 26.22.2002, http://www.egov.vic.gov.au/Documents/DCforMPeg7.doc

[12] XML and MPEG-7 for Interactive Annotation and Retrieval using Semantic Metadata, Lux Mathias et al, J.UCS Vol 8 Issue 10, 28.10.2002, http://www.jucs.org/jucs_8_10/xml_and_mpeg_7